Shrinkage estimators of intercept parameters of two simple regression models with suspected equal slopes

Shahjahan Khan
Department of Mathematics & Computing
Australian Centre for Sustainable Catchments
University of Southern Queensland
Toowoomba, Queensland, Australia
Email: khans@usq.edu.au

Abstract

Estimators of the intercept parameter of a simple linear regression model involves the slope estimator. In this paper, we consider the estimation of the intercept parameters of two linear regression models with normal errors, when it is *apriori* suspected that the two regression lines are parallel, but in doubt. We also introduce a *coefficient of distrust* as a measure of degree of lack of trust on the uncertain prior information regarding the equality of two slopes. Three different estimators of the intercept parameters are defined by using the sample data, the non-sample uncertain prior information, an appropriate test statistic, and the coefficient of distrust. The relative performances of the unrestricted, shrinkage restricted and shrinkage preliminary test estimators are investigated based on the analyses of the bias and risk functions under quadratic loss. If the prior information is precise and the coefficient of distrust is small the shrinkage preliminary test estimator over performs the other estimators. An example based on a medical study is used to illustrate the method.

Keywords: Central and non-central F-distribution, coefficient of distrust, non-sample uncertain prior information, parallel regression lines, quadratic bias and risk, shrinkage restricted and preliminary test estimators.

AMS 2000 Subject Classification: Primary 62F30 and Secondary 62J05.

1 Introduction

The problem of suspected parallelism arises in many bioassays and studies in the areas of social as well as physical sciences. When sample data of two categories of respondents are available on the same response and explanatory variables the data can be modelled by two separate regression lines. Experts in the field, based on the knowledge of the subject or previous experience, may suspect that the slopes of the two regression lines are equal. Such a non-sample prior information about the value of the slopes can be represented by a null hypothesis. Often researchers have varying degrees of trust on such a non-sample prior information, and are able to express the *coefficient of distrust*,

$0 \leq \ d \leq 1$, as a measure of degree of lack of trust on the null hypothesis (cf Khan and Saleh, 2001). The additional uncertain non-sample prior information such as the null hypothesis of equality of the slopes and the coefficient of distrust along with the sample data are used to define various estimators with a view to improving the statistical properties of the estimators.

Consider a clinical/medical study where the experimenter has collected two different data sets on the effect of two drugs for building two separate regression models. Alternatively, consider a sociologist or psychologist who has constructed two regression equations, one set for the males and another for the females. In both cases it may be useful to get some insight into whether or not the parameters of the two different regression models differ significantly across the two data sets. Moreover, the researcher may wish to combine the two data sets to formulate an overall regression model, if the respective parameters of the two different regression models do not differ significantly. However, in practical problems, the parameters of the models are usually unknown and the equality of slopes can only be suspected. This kind of suspicion may be treated as non-sample *uncertain prior information* and can be incorporated in the estimation of the parameters of the models.

Customarily, the regression parameters are estimated by using the sample data alone. However, it is well known that the inclusion of non-sample prior information in the estimation of parameters is likely to improve the quality of the estimator in terms of desirable statistical properties. Bancroft (1944) first introduced the idea of preliminary test estimator. Such an estimator uses both the sample data and non-sample prior information in the form of a suspected null hypothesis. Appropriate statistical test is performed to remove the element of doubt in the null hypothesis. Then the preliminary test estimator is defined as a function of the sample data, the non-sample prior information and the test statistic. Khan and Saleh (2001) introduced the idea of using the coefficient of distrust in the estimation of parameters. The same idea can be applied to the parallelism problem with two regression equations, when it is apriori suspected that the slopes of the two regression lines are equal, but in doubt. Khan (2003) has adopted this approach to estimate the slope parameters of two suspected parallel regression models. In this paper we define and investigate three different estimators of the intercept parameters of two linear regression lines by using the sample data, the non-sample uncertain prior information, appropriate test statistic as well as the coefficient of distrust. The properties of the three different estimators are investigated through detailed analysis of the bias and quadratic risk functions.

Data for two regression equations can be expressed as

$$y_{1j} = \theta_1 + \beta_1 x_{1j} + \epsilon_{1j};\ j = 1, 2, \cdots, n_1 \text{ and } y_{2j} = \theta_2 + \beta_2 x_{2j} + \epsilon_{2j}; j = 1, 2, \cdots, n_2 \quad (1.1)$$

where $\boldsymbol{y} = [\boldsymbol{y}_1', \ \boldsymbol{y}_2']$ in which $\boldsymbol{y}_1 = [y_{11}, y_{12}, \cdots, y_{1n_1}]'$, $\boldsymbol{y}_2 = [y_{21}, y_{22}, \cdots, y_{2n_2}]'$, and $\boldsymbol{x} = [\boldsymbol{x}_1', \ \boldsymbol{x}_2']$ in which $\boldsymbol{x}_1 = [x_{11}, x_{12}, \cdots, x_{1n_1}]'$ and $\boldsymbol{x}_2 = [x_{21}, x_{22}, \cdots, x_{2n_2}]'$. Note that y_{ij} is the j^{th} response of the i^{th} model and ϵ_{ij} is the associated error component; x_{ij} is the j^{th} value of the predictor variable in the i^{th} model; and β_i and θ_i are the slope and intercept parameters of the i^{th} regression equation for $i = 1, 2$. We assume that the errors are identically and independently distributed as normal variables with mean 0 and unknown variance σ^2. Our problem is to estimate the vector of intercept parameters, $\boldsymbol{\theta} = (\theta_1, \theta_2)'$, and that of the slope parameters, $\boldsymbol{\beta} = (\beta_1, \beta_2)'$, when equality of slopes is suspected, but in doubt.

The two regression equations can be combined in a single model as

$$\boldsymbol{y} = X\boldsymbol{\Phi} + \boldsymbol{e} \tag{1.2}$$

where $\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \\ \boldsymbol{y}_2 \end{pmatrix}$, $X = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \boldsymbol{x}_1 & \mathbf{0} \\ \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \boldsymbol{x}_2 \end{pmatrix}$, $\boldsymbol{\Phi} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \beta_1 \\ \beta_2 \end{pmatrix}$ and $\boldsymbol{e} = \begin{pmatrix} \boldsymbol{e}_1 \\ \\ \boldsymbol{e}_2 \end{pmatrix}$. Now, if it is suspected that the two lines are parallel then the suspicion in the form of non-sample *uncertain prior information*, say β, is expressed by the null hypothesis

$$H_0 : \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \boldsymbol{\Phi} = \begin{pmatrix} \beta \\ \beta \end{pmatrix}. \tag{1.3}$$

In general, the null hypothesis of equality of slopes is given by $H_0 : \ C\boldsymbol{\Phi} = \boldsymbol{r}$, and the alternative hypothesis, H_a : negation of the H_0, where C is a matrix and $\boldsymbol{\Phi}$ and $\boldsymbol{r}$ are vectors of appropriate orders. It is under the general null hypothesis in (1.3), we wish to estimate the intercept parameters of the regression lines represented in (1.1).

The problem under consideration falls in the realm of statistical problems known as inference in the presence of *uncertain prior information.* The usual practice in the literature is to treat such *uncertain prior information* specified by H_0 as a "nuisance parameter". Then the uncertainty in the form of the "nuisance parameter" is removed by 'testing it out'. In a series of papers Bancroft (1944, 1964, 1972) addressed the problem, and proposed the well known *preliminary test* estimator. A host of other authors, notably Kitagawa (1963), Han and Bancroft (1968), Saleh and Han (1990), Ali and Saleh (1990), and Mahdi et al. (1998) contributed in the development of the method under the normal theory. Furthermore, Saleh and Sen (e.g., 1978, 1985) published a series of articles in this area exploring the nonparametric as well as the asymptotic theory based on the least square estimators. Bhoj and Ahsanullah (1993, 1994) discussed the problem of estimation of conditional mean for simple regression model. Khan and Saleh (1997) discussed the problem of shrinkage pre-test estimation for the multivariate Student-t regression model.

In this paper, we define the maximum likelihood estimator (mle) of the elements of $\boldsymbol{\Phi}$ in (1.2) assuming that the errors are independently and identically distributed as normal variables with mean 0 and unknown variance σ^2. Such an estimator is known as the *unrestricted estimator* (UE). Then we define the *shrinkage restricted estimator* (SRE) of $\boldsymbol{\theta}$ under the constraint of the H_0 and by using the coefficient of distrust $0 \leq d \leq 1$ as a measure of the degree of lack of trust on the H_0. Finally, we define the *shrinakge preliminary test estimator* (SPTE) of $\boldsymbol{\theta}$ by using an appropriate test statistic that can be employed to test the null hypothesis in addition to the sample and non-sample prior information. The main objective of the paper is to study the properties of the three different estimators, namely the UE, SRE and SPTE, for the intercept parameters of the two suspected parallel regression lines. Also, we investigate the relative performances of the estimators under different conditions. The analysis of the performances of the estimators are provided that can be used as a basis to select a 'best' estimator in a given situation. The comparisons of the estimators are based on the criteria of unbiasedness and risk under quadratic loss, both analytically and graphically.

The *preliminary test* estimator (PTE) is defined as a function of the test statistic appropriate for testing the null hypothesis as well as the UE and RE. In fact, it is an extrem choice between the UE and RE. The *shrinkage preliminary test* estimator (SPTE) is defined as a function of the test statistic appropriate for testing the null hypothesis, the UE and SRE. The later depends on the coefficient of distrust. From the definition, the SPTE yields the UE if the value of the coefficient of distrust is $d = 1$, regardless of the acceptance of or rejection of H_0 at any level of significance. On the other hand the SPTE becomes the SRE if the H_0 is not rejected at any given level of significance and the value of $d \neq 1$. When $d = 1$ the SRE becomes UE. Therefore, the shrinkage preliminary test estimator indeed gives us a compromising choice between the two estimators, UE and SRE except for $d = 0$ or 1. Although PTE is an extreme choice between the UE and RE, the SPTE allows a compromise between the UE and SRE. Such a smooth compromise between the two extremes, UE and RE, has been discussed by Khan and Saleh (1995).

In the next section, we define three different estimators of the previously defined vector of the intercept parameters. Some important results, that are necessary for the computations of bias and quadratic risk of the estimators are discussed in section 3. The expressions for bias of the estimators and their analyses are provided in section 4. The performance comparison of the estimators of the intercept parameters based on the quadratic risk criterion is discussed in section 5. Section 6 provides an example based on a set of clinical data. Some concluding remarks are included in section 7.

2 Formulation of the estimators

Assume that the error ϵ_{ij} in (1.1) is independent and identically distributed as a normal variable with $\mathrm{E}(\epsilon_{ij}) = 0$ and $\mathrm{Var}(\epsilon_{ij}) = \sigma^2$ (unknown) for $i = 1, 2$ and all j. Then the *unrestricted estimator* (UE) of β_i and θ_i are obtained by applying the method of maximum likelihood (or equivalently the least squares method) as

$$\tilde{\beta}_i = \sum_{j=1}^{n_i} \frac{(x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i)}{n_i Q_i}, \quad \tilde{\theta}_i = \bar{y}_i - \tilde{\beta}_i \bar{x}_i \tag{2.1}$$

where $\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}, \quad \bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$ and $n_i Q_i = \sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2$ for $i = 1, 2$. Thus the *unrestricted estimator* (UE) of the vectors of the slope and intercept, $\boldsymbol{\beta} = (\beta_1,\ \beta_2)'$ and $\boldsymbol{\theta} = (\theta_1,\ \theta_2)'$ becomes

$$\tilde{\boldsymbol{\beta}} = (\tilde{\beta}_1,\ \tilde{\beta}_2)', \quad \tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1,\ \tilde{\theta}_2)' = \bar{\boldsymbol{y}} - \boldsymbol{T}\tilde{\boldsymbol{\beta}} \tag{2.2}$$

where $\bar{\boldsymbol{y}} = (\bar{y}_1,\ \bar{y}_2)'$ and $\boldsymbol{T} = \mathrm{Diag}\{\bar{x}_1,\ \bar{x}_2\}$, a 2×2 diagonal matrix. When the null hypothesis of equality of slopes holds, then the *restricted estimator* (RE) of the slope parameter becomes

$$\hat{\beta} = \frac{1}{nQ}\sum_{i=1}^{2} n_i Q_i \tilde{\beta}_i \quad \text{with} \quad nQ = \sum_{i=1}^{2} n_i Q_i \text{ and } n = n_1 + n_2. \tag{2.3}$$

Here $\hat{\beta}$ is the maximum likelihood estimator of the slope when the null hypothesis is true. Therefore, the *restricted estimator* (RE) of the vectors $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are defined as

$$\hat{\boldsymbol{\beta}} = \hat{\beta}\boldsymbol{l}_2 = (\hat{\beta}, \hat{\beta})', \quad \hat{\boldsymbol{\theta}} = \bar{\boldsymbol{y}} - \boldsymbol{T}\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\theta}} + \boldsymbol{T}J\tilde{\boldsymbol{\beta}} \tag{2.4}$$

where $J = I_2 - \frac{\boldsymbol{l}_2\boldsymbol{l}_2'}{nQ}D_2^{-1}$ in which $D_2^{-1} = \mathrm{Diag}\{n_1Q_1, n_2Q_2\}$, $\boldsymbol{l}_2$ is a 2-tuples of ones and I_2 is the identity matrix of order 2.

Introducing the coefficient of distrust on the H_0, the *shrinkage restricted estimator* (SRE) of the vector $\boldsymbol{\theta}$ is defined as

$$\hat{\boldsymbol{\theta}}_d = d\tilde{\boldsymbol{\theta}} + (1-d)\hat{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}} + (1-d)\boldsymbol{T}J\tilde{\boldsymbol{\beta}} \tag{2.5}$$

The *uncertainty* in the null hypothesis H_0 is removed by using an appropriate test statistic. For the current problem, we consider the likelihood ratio test given by the following statistic

$$L_n = \frac{(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})' D_3^{-1} (\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})}{s^2} \tag{2.6}$$

where $D_3^{-1} = \mathrm{Diag}\{\frac{1}{n_1Q_1} + \frac{1}{nQ}, \frac{1}{n_2Q_2} + \frac{1}{nQ}\}$ and $s^2 = \frac{1}{m}\sum_{i=1}^{2}\sum_{j=1}^{n_i}[(y_{ij} - \bar{y}_i) - \tilde{\beta}_i(x_{ij} - \bar{x}_i)]^2$ with $m = (n-4)$, and the numerator can be expressed as

$$(\tilde{\beta}_1 - \hat{\beta})^2 \frac{(n_1Q_1nQ)}{(n_1Q_1 + nQ)} + (\tilde{\beta}_2 - \hat{\beta})^2 \frac{(n_2Q_2nQ)}{(n_2Q_2 + nQ)}. \tag{2.7}$$

Under the null hypothesis, the above test statistic follows a central F-distribution with 2 and m degrees of freedom. Let F_α denote the $(1-\alpha)^{th}$ quantile of an $F_{2,m}$ variable such that $(1-\alpha)\times 100\%$ area under the curve of the distribution is to the left of F_α. Then, the *preliminary test estimator* (PTE) of $\boldsymbol{\theta}$ is defined as

$$\hat{\boldsymbol{\theta}}^{pt} = \tilde{\boldsymbol{\theta}} - (\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})I(L_n < F_\alpha) = \tilde{\boldsymbol{\theta}} + \boldsymbol{T}J\tilde{\boldsymbol{\beta}}I(L_n < F_\alpha) \tag{2.8}$$

where $I(A)$ denotes an indicator function of the set A. The PTE, defined above, is a convex combination of the UE and RE, and depends on the random coefficient, $\zeta = I(L_n < F_\alpha)$ whose value is 1 when the null hypothesis is accepted and 0 otherwise. Also note that the PTE is an extreme compromise between the UE and RE. At a given level of significance, the PTE may simply be either the UE or RE depending on the rejection and acceptance of the null hypothesis respectively. Therefore, for large values of L_n the PTE becomes the UE and for smaller values of L_n the PTE turns out to be the RE. Obviously, the PTE is a function of the test statistic as well as the level of significance, α. Hence, the PTE may change its value with a change in the choice of α. Therefore, a search for an optimal value of α may be desirable. In this paper, the optimality of the level of significance is in the sense of minimising the maximum risk of an estimator. Methods are available in the literature that provide optimal α, (see Akaike (1972), for instance). Another fact about the PTE is that it does not allow smooth transition between the two extremes, the UE and RE. Khan and Saleh (1995) provided a *shrinkage preliminary test estimator* to overcome such a problem.

Now we define the *shrinkage preliminary test estimator* (SPTE) of the intercept vector, $\boldsymbol{\theta}$ as follows

$$\begin{aligned} \hat{\boldsymbol{\theta}}_d^{pt} &= \hat{\boldsymbol{\theta}}_d I(L_n < F_\alpha) + \tilde{\boldsymbol{\theta}} I(L_n \geq F_\alpha) \\ &= \tilde{\boldsymbol{\theta}} + (1-d)\boldsymbol{T}J\tilde{\boldsymbol{\beta}}I(L_n < F_\alpha). \end{aligned} \tag{2.9}$$

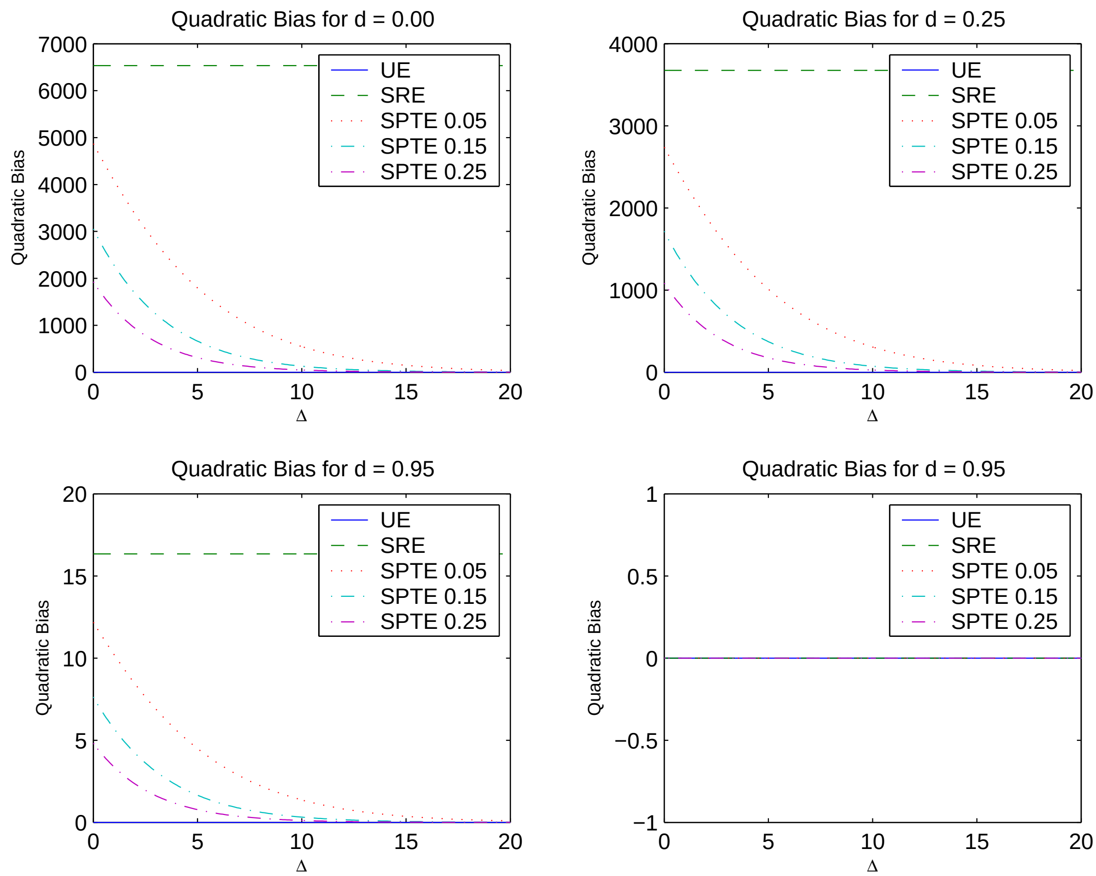
From the definition the SPTE becomes the UE when $d = 1$ and the PTE when $d = 0$.

Since we have defined three different estimators for the slope and the intercept parameter, a natural question arises as to which estimator should be used, and why? The answer to the question requires to investigate the performances of the estimators under different conditions. To study the properties of the above estimators of the intercept vector, some essential results are provided in the next section.

3 Some Preliminaries

In this section, we provide some useful results that are instrumental to the computation of expressions for bias and risk under quadratic loss function for the three different estimators. First, observe that the joint distribution of $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\theta}}$ is multivariate normal with

$$\mathrm{E}\begin{pmatrix}\tilde{\boldsymbol{\theta}}\\ \tilde{\boldsymbol{\beta}}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\theta}\\ \boldsymbol{\beta}\end{pmatrix} \text{ and covariance matrix, } \mathrm{Cov}\begin{pmatrix}\tilde{\boldsymbol{\theta}}\\ \tilde{\boldsymbol{\beta}}\end{pmatrix} = \sigma^2\begin{pmatrix} D_1 & D_{12} \\ D_{21} & D_2\end{pmatrix} \tag{3.1}$$



h

Figure 1: **Quadratic bias of the estimators for selected values of d.**

where $D_{12} = D_{21}' = -D_2\boldsymbol{T}$ and $D_1 = \frac{\mathrm{Cov}(\tilde{\boldsymbol{\theta}})}{\sigma^2} = \boldsymbol{\psi} + \boldsymbol{T}D_2\boldsymbol{T}'$ with $\boldsymbol{\psi} = \mathrm{Diag}\left\{\frac{1}{n_1}, \frac{1}{n_2}\right\}$.

Note that the matrix D_2 has been specified in the definition of J in equation (2.4). Also note that $JD_2J' = D_2$, $D_2J'\boldsymbol{T} = -D_{12}\boldsymbol{T} + \frac{\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}'}{nQ}$ with $\bar{\boldsymbol{x}} = (\bar{x}_1, \bar{x}_2)'$. Moreover, the joint distribution of the elements of $\hat{\boldsymbol{\beta}}$ is bivariate normal with the mean vector,

$$\mathrm{E}[\hat{\boldsymbol{\beta}}] \ = \boldsymbol{\beta}_0 = \beta\boldsymbol{l}_2 \text{ and covariance matrix, } \mathrm{Cov}[\hat{\boldsymbol{\beta}}] = \sigma^2 D_2^* \tag{3.2}$$

where $\boldsymbol{l}_2$ is a vector of 1's of order two and $D_2^* = \mathrm{Diag}\{\frac{1}{nQ}, \ \frac{1}{nQ}\}$. Finally the distribution of $(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})$ is bivariate normal with

$$\mathrm{E}[\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}] = \boldsymbol{\delta} \ \text{ and covariance matrix, } \mathrm{Cov}[\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}] \ = \sigma^2 D_3 \tag{3.3}$$

where $\boldsymbol{\delta} = (\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ and $D_3 = D_2 + D_2^* = \mathrm{Diag}\{\frac{1}{n_1Q_1} + \frac{1}{nQ}, \ \frac{1}{n_2Q_2} + \frac{1}{nQ}\}$.

In the next section, we derive the expressions of bias for the previously defined estimators of the intercept parameters.

4 The bias of estimators

First, the expression for the bias of UE of $\boldsymbol{\theta}$ is obtained as

$$B_1(\tilde{\boldsymbol{\theta}}) = E(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathbf{0}. \tag{4.1}$$

Thus $\tilde{\boldsymbol{\theta}}$ is an unbiased estimator of $\boldsymbol{\theta}$. This is a well-known property of the mle for normal models. The bias of the RE of $\boldsymbol{\theta}$ is found to be

$$B_2^*(\hat{\boldsymbol{\theta}}) = E(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \boldsymbol{T}\boldsymbol{\delta} \tag{4.2}$$

where $\boldsymbol{\delta} = J\boldsymbol{\beta} = \boldsymbol{\beta} - \beta\boldsymbol{l}_2$, deviation of $\boldsymbol{\beta}$ from its suspected value under H_0. Clearly, the RE is biased when $\boldsymbol{T} \neq \mathbf{0}$. The amount of bias becomes unbounded as $\boldsymbol{\delta} \to \infty$, that is, if the true value of $\boldsymbol{\beta}$ is far away from it's hypothesized value, $\beta\boldsymbol{l}_2$. On the other hand the bias is zero when the null hypothesis is true. Thus unlike the UE, the RE is biased under the alternative hypotheis.

The bias of the SRE of $\boldsymbol{\theta}$ is found to be

$$B_2(\hat{\boldsymbol{\theta}}_d) = E(\hat{\boldsymbol{\theta}}_d - \boldsymbol{\theta}) = (1-d)\boldsymbol{T}\boldsymbol{\delta} \tag{4.3}$$

So, the SRE is biased, if either $d \neq 1$, or $\boldsymbol{T} \neq \mathbf{0}$, or $\boldsymbol{\delta}$ is non-zero. Thus for $d \neq 1$ and $\boldsymbol{T} \neq \mathbf{0}$, the amount of bias becomes unbounded as $\boldsymbol{\delta} \to \infty$. Thus unlike the UE, the SRE is biased, and assumes its largest value at $d = 0$. Clearly, $B_2(\hat{\boldsymbol{\theta}}_d) \leq B_2^*(\hat{\boldsymbol{\theta}})$.

The bias expression for the PTE is obtained as

$$B_3^*(\hat{\boldsymbol{\theta}}^{pt}) = E(\hat{\boldsymbol{\theta}}^{pt} - \boldsymbol{\theta}) = \boldsymbol{T}\boldsymbol{\delta} G_{3,m}(l_\alpha; \Delta) \tag{4.4}$$

where $\Delta = \frac{\boldsymbol{\delta}' D_2^{-1} \boldsymbol{\delta}}{\sigma^2}$, $l_\alpha = \frac{1}{3} F_\alpha$ and $G_{3,m}(l_\alpha; \Delta) = \int_{z=0}^{l_\alpha} f(z) dz$ in which Z has a non-central $F-$distribution. For the computational purposes, $G_{3,m}(l_\alpha; \Delta)$ can be written as

$$G_{3,m}(l_\alpha; \Delta) = \sum_{r=0}^{\infty} \frac{e^{-\frac{\Delta}{2}} (\frac{\Delta}{2})^r}{r!} IB_{q_\alpha}^1 \left(\frac{3}{2} + r, \frac{m}{2}\right) \tag{4.5}$$

where $IB_{q_\alpha}^1 \left(\frac{3}{2} + r, \frac{m}{2}\right)$ is the incomplete beta function ratio and $q_\alpha = \frac{m}{m + F_{1,m}(\alpha)}$. In the derivation of the bias expression for the PTE we use the result of Appendix B1 of Judge and Bock (1978) as well as the results in the previous section.

Obviously, the PTE is a biased estimator, and the amount of bias depends on the value of $G_{3,m}(\cdot)$, the cdf of a non-central F distribution, and the extent of departure of the parameter from its value under null hypothesis. However, since $0 \leq G_{3,m}(\cdot) \leq 1$, the bias of the PTE is always smaller than that of the RE, if $\Delta \neq 0$. So, in general $B_3^*(\hat{\boldsymbol{\theta}}^{pt}) \leq B_2(\hat{\boldsymbol{\theta}}_d) \leq B_2^*(\hat{\boldsymbol{\theta}})$.

Finally, the bias expression for the SPTE is obtained as

$$B_3(\hat{\boldsymbol{\theta}}^{pt}) = E(\hat{\boldsymbol{\theta}}^{pt} - \boldsymbol{\theta}) = (1-d)\boldsymbol{T}\boldsymbol{\delta} G_{3,m}(l_\alpha; \Delta). \tag{4.6}$$

Like the PTE, the SPTE is a biased estimator, except for $d = 1$, or $\boldsymbol{T} = \mathbf{0}$ or $\boldsymbol{\delta} = \mathbf{0}$. The amount of bias depends on the value of $G_{3,m}(\cdot)$. The bias of the SPTE becomes the same as that of the PTE if $d = 0$. However, the bias of the SPTE is always less than or equal to that of the SRE for all values of d. Therefore, $B_3(\hat{\boldsymbol{\theta}}_d^{pt}) \leq B_3^*(\hat{\boldsymbol{\theta}}^{pt}) \leq B_2(\hat{\boldsymbol{\theta}}_d) \leq B_2^*(\hat{\boldsymbol{\theta}})$.

The Graph of Quadratic Bias:

The bias function of the intercept vector is also a vector of the same order. So direct comparison of the bias functions of the estimators are not meaningful. To compare the overall bias of the estimators we define the quadratic bias as the vector product of the bias by itself. The quadratic bias is a scalar and it can be compared across the estimators. The plot of the quadratic bias function of the UE, SRE and SPTE with $\alpha = 0.05,\ 0.15$ and 0.25 are provided in Figure 1 for different values of the non-centrality parameter Δ and $\sigma = 1$. As expected, the quadratic bias of the UE is 0 for all values of Δ and that of the SRE is unbounded and increases as the value of Δ grows large. The quadratic bias of the SPTE is a function of the level of significance. As shown in the bottom two graphs in Figure 1, the shape of the curve of the quadratic bias function of the SPTE is skewed to the right. Very near at $\Delta = 0$ it starts at the largest value and moves downward sharply and then gradually declines to the horizontal axis. The quadratic bias of the SPTE increases as the preselected level of significance decreases. This is quite clear from the first three graphs in Figure 1. The quadratic bias function of the SRE and SPTE increases as the variance of the population becomes larger.

Smaller the value of d higher in the value of quadratic bias. For $d = 1$ the quadratic bias for all estimators is 0. Also for very large Δ the quadratic bias approaches 0. For increased values of σ the shape of the quadratic bias function remains the same but the magnitude increases.

The graphs in Figure 1 are produced for the quadratic bias functions of the intercept parameters for selected values of d and $\sigma = 1$. Similar graphs for the quadratic bias functions can also be produced for different values of σ.

5 The risk of estimators

The quadratic error loss function of an estimator, $\boldsymbol{t}^*$ to estimate the parameter, $\boldsymbol{\mu}$, is defined to be

$$L(\boldsymbol{t}^*, W_1, \boldsymbol{\mu}) = (\boldsymbol{t}^* - \boldsymbol{\mu})' W_1 (\boldsymbol{t}^* - \boldsymbol{\mu})$$

where W_1 is a positive definite matrix of appropriate dimension. Then the quadratic risk of $\boldsymbol{t}^*$ in estimating $\boldsymbol{\mu}$ is the expected value of $L(\boldsymbol{t}^*, W_1, \boldsymbol{\mu})$. Thus for the intercept vector, $\boldsymbol{\theta}$, the quadratic risk function is given by

$$R(\boldsymbol{\theta}^*, W_1, \boldsymbol{\theta}) = \mathrm{E}(\boldsymbol{\theta}^* - \boldsymbol{\theta})' W_1 (\boldsymbol{\theta}^* - \boldsymbol{\theta}) \tag{5.1}$$

where $\boldsymbol{\theta}^*$ is an estimator of $\boldsymbol{\theta}$ and W_1 is a positive definite matrix of appropriate dimension. Therefore, the expression of the quadratic risk for the UE of $\boldsymbol{\theta}$ is obtained as

$$R_1(\tilde{\boldsymbol{\theta}}; W_1) = E(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})' W_1 (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \sigma^2 tr(W_1 D_1). \tag{5.2}$$

Similarly, the risk of the RE of $\boldsymbol{\theta}$ is found to be

$$R_2(\hat{\boldsymbol{\theta}}; W_1) = E(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' W_1 (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \sigma^2 tr(W_1 D_1) + \boldsymbol{\delta}' \boldsymbol{T}' W_1 \boldsymbol{T} \boldsymbol{\delta}. \tag{5.3}$$

Now, the quadratic risk expression of the PTE is given by

$$\begin{aligned} R_3^*(\hat{\boldsymbol{\theta}}^{pt}; W_1) &= E(\hat{\boldsymbol{\theta}}^{pt} - \boldsymbol{\theta})' W_1 (\hat{\boldsymbol{\theta}}^{pt} - \boldsymbol{\theta}) = \sigma^2 tr(W_1 D_1) \Big\{ 1 - G_{3,m}(l_\alpha; \Delta) \Big\} \\ & \qquad + \boldsymbol{\delta}' \boldsymbol{T}' W_1 \boldsymbol{T} \boldsymbol{\delta} \Big\{ 2 G_{3,m}(l_\alpha; \Delta) - G_{5,m}(l_\alpha^*; \Delta) \Big\}. \end{aligned} \tag{5.4}$$

The proof of the above results is straight forward by using the Appendix B1 of Judge and Bock (1978).

Finally the quadratic risk function of the SPTE is found to be

$$\begin{aligned} R_3(\hat{\boldsymbol{\theta}}_d^{pt}; W_1) &= E(\hat{\boldsymbol{\theta}}_d^{pt} - \boldsymbol{\theta})' W_1 (\hat{\boldsymbol{\theta}}_d^{pt} - \boldsymbol{\theta}) \\ &= \sigma^2 \Big\{ tr(W_1 D_1) - (1 - d^2) tr(W_1 D_3) G_{3,m}(l_\alpha; \Delta) \Big\} + \boldsymbol{\delta}' \boldsymbol{T}' W_1 \boldsymbol{T} \boldsymbol{\delta} \\ & \quad \times \Big\{ 2(1-d) G_{3,m}(l_\alpha; \Delta) - (1 - d^2) G_{5,m}(l_\alpha^*; \Delta) \Big\}. \end{aligned} \tag{5.5}$$

When there is no distrust on the null hypothesis, that is $d = 0$, then $R_3(\hat{\boldsymbol{\theta}}_d^{pt}; W_1) = R_3^*(\hat{\boldsymbol{\theta}}^{pt}; W_1)$; and when there is total distrust on the null hypothesis, that is $d = 1$, we get $R_3(\hat{\boldsymbol{\theta}}_d^{pt}; W_1) = R_1(\tilde{\boldsymbol{\theta}}; W_1)$. Thus the quadratic risk of the SPTE yields that of the PTE and UE for the two extreme values of d.

5.1 Risk analysis for estimators of intercept

In this section, we compare the performance of the estimators of the intercept parameter vector based on the quadratic risk criterion.

Comparison of UE and SRE

First consider the difference between the risks of the UE and SRE,

$$H_{12}(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}; W_1) = R_1(\tilde{\boldsymbol{\theta}}; W_1) - R_2(\hat{\boldsymbol{\theta}}_d; W_1) = -(1-d)^2 \boldsymbol{\delta}' \boldsymbol{T}' W_1 \boldsymbol{T} \boldsymbol{\delta} = -(1-d)^2 \sigma^2 \Delta_T \tag{5.6}$$

where $\Delta_T = \frac{\boldsymbol{\delta}' \boldsymbol{T}' W_1 \boldsymbol{T} \boldsymbol{\delta}}{\sigma^2}$. Thus the value of $H_{12}(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}; W_1)$ is negative, zero or positive depending on

$$(1-d)^2 \sigma^2 \Delta_T \gtreqless 0, \text{ or } \Delta_T \gtreqless 0 \quad \text{when } d \neq 1. \tag{5.7}$$

Note that when $d \neq 1$, the UE dominates the SRE in terms of having smaller risk. However, for $d = 1$, the UE and SRE have the same risk. Nevertheless, it is very unlikely that d will be 1, or even near 1. Furthermore, the SRE has larger risk than the UE if $d = 0$.

Comparison of UE and SPTE

The risk-difference of the UE and the SPTE is given by

$$H_{13}(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}_d^{pt}; W_1) = R_1(\tilde{\boldsymbol{\theta}}; W_1) - R_3(\hat{\boldsymbol{\theta}}_d^{pt}; W_1) = (1-d)^2\sigma^2 tr(W_1 D_3) G_{3,m}(l_\alpha; \Delta)$$
$$-\boldsymbol{\delta}'\boldsymbol{T}'W_1\boldsymbol{T}\boldsymbol{\delta}\Big\{2(1-d)G_{3,m}(l_\alpha; \Delta) - (1-d^2)G_{5,m}(l_\alpha^*; \Delta)\Big\}. \quad (5.8)$$

Thus we have

$$H_{13}(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}_d^{pt}; W_1) \gtreqless 0 \text{ whenever } \Delta_T \lesseqgtr \frac{(1-d)^2 tr(W_1 D_3) G_{3,m}(l_\alpha; \Delta)}{\Big\{2(1-d)G_{3,m}(l_\alpha; \Delta) - (1-d^2)G_{5,m}(l_\alpha^*; \Delta)\Big\}}. \quad (5.9)$$

In a special case, when $W_1 = D_3^{-1}$ then (5.9) becomes

$$\frac{\boldsymbol{\delta}'\boldsymbol{T}'D_3^{-1}\boldsymbol{T}\boldsymbol{\delta}}{\sigma^2} \lesseqgtr \frac{2(1-d)^2 G_{3,m}(l_\alpha; \Delta)}{\Big\{2(1-d)G_{3,m}(l_\alpha; \Delta) - (1-d^2)G_{5,m}(l_\alpha^*; \Delta)\Big\}}. \quad (5.10)$$

In another special case, when $d = 1$, $H_{13}(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}_d^{pt}; W_1) = 0$, and hence the risk of the UE equals that of the SPTE if there is total distrust on the uncertain prior information. Furthermore, for $d = 0$, we get

$$H_{13}(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}_d^{pt}; W_1) = \sigma^2 tr(W_1 D_3) G_{3,m}(l_\alpha; \Delta) - \boldsymbol{\delta}'\boldsymbol{T}'D_3^{-1}\boldsymbol{T}\boldsymbol{\delta}\Big\{2G_{3,m}(l_\alpha; \Delta) - G_{5,m}(l_\alpha^*; \Delta)\Big\}. \quad (5.11)$$

So

$$H_{13}(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}_d^{pt}; W_1) \gtreqless 0 \text{ whenever } \Delta_T \lesseqgtr \frac{tr(W_1 D_3) G_{3,m}(l_\alpha; \Delta)}{\Big\{2G_{3,m}(l_\alpha; \Delta) - (1-d^2)G_{5,m}(l_\alpha^*; \Delta)\Big\}}. \quad (5.12)$$

Thus, the SPTE over performs the UE if $\Delta_T < \frac{tr(W_1 D_3) G_{3,m}(l_\alpha; \Delta)}{\Big\{2G_{3,m}(l_\alpha; \Delta) - (1-d^2)G_{5,m}(l_\alpha^*; \Delta)\Big\}}$.

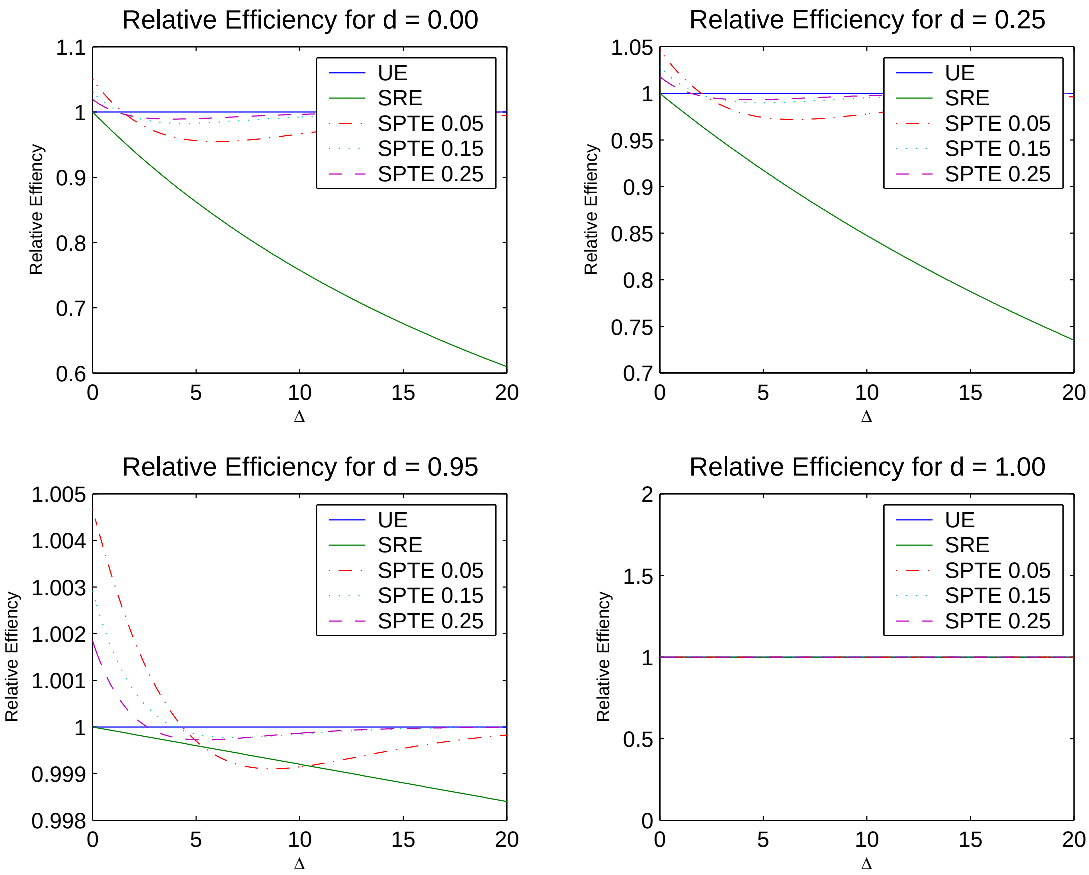
Comparison of SPTE and SRE

The difference between the risks of the SPTE and SRE is

$$H_{32}(\hat{\boldsymbol{\theta}}_d^{pt}, \hat{\boldsymbol{\theta}}_d; W_1) = R_3(\hat{\boldsymbol{\theta}}_d^{pt}; W_1) - R_2(\hat{\boldsymbol{\theta}}_d; W_1)$$
$$= -(1-d)^2\sigma^2 tr(W_1 D_3) G_{3,m}(l_\alpha; \Delta) - \boldsymbol{\delta}'\boldsymbol{T}'W_1\boldsymbol{T}\boldsymbol{\delta}$$
$$\times\Big\{(1-d)^2 - 2(1-d)G_{3,m}(l_\alpha; \Delta) - (1-d^2)G_{5,m}(l_\alpha^*; \Delta)\Big\}. \quad (5.13)$$

Now, from (5.13) we get $H_{32}(\hat{\boldsymbol{\theta}}_d^{pt}, \hat{\boldsymbol{\theta}}_d; W_1) \lesseqgtr 0$ according as

$$\frac{\boldsymbol{\delta}'\boldsymbol{T}'W_1\boldsymbol{T}\boldsymbol{\delta}}{\sigma^2} \gtreqless \frac{(1-d) tr(W_1 D_3) G_{3,m}(l_\alpha; \Delta)}{2(1-d)G_{3,m}(l_\alpha; \Delta) - (1-d^2)G_{5,m}(l_\alpha^*; \Delta) - (1-d)^2}. \quad (5.14)$$



h

Figure 2: **Relative efficiency of the estimators for selected values of d.**

Therefore, based on (5.14), SRE performs better than the SPTE if

$$\Delta_T < \frac{(1-d)tr(W_1 D_3)G_{3,m}(l_\alpha;\Delta)}{2G_{3,m}(l_\alpha;\Delta) - (1+d)G_{5,m}(l_\alpha^*;\Delta) - (1-d)} \tag{5.15}$$

and the SPTE dominates over the SRE whenever

$$\Delta_T > \frac{(1-d)tr(W_1 D_3)G_{3,m}(l_\alpha;\Delta)}{2G_{3,m}(l_\alpha;\Delta) - (1+d)G_{5,m}(l_\alpha^*;\Delta) - (1-d)}. \tag{5.16}$$

To compare the relative performances of the estimators based on the quadratic risk criterion, we define the relative efficiency of the estimators from the quadratic risk functions.

The graphs in Figure 2 are produced for the relative efficiency functions of the intercept parameters for selected values of d and $\sigma = 1$. Similar graphs for the relative efficiency functions can also be produced for different values of σ. Graphs for the quadratic risk functions can also be produced in the same fashion.

For $d = 1$ all three estimators are equally efficient. But from the premises of the choice of d is is unlikely to be even close to 1. The relative efficiency of the SRE relative to the UE is always less than 1 for all σ and $d \neq 1$. Similarly, The relative efficiency of the SPTE relative to the UE is larger than that of the SRE, except at $d = 1$. At $\Delta = 0$, the relative efficiency of the SPTE attains its maximum, and then declines to the 1-line for some moderate values of Δ. For a small range of values of Δ, the relative efficiency of the SPTE is less than 1, but as $\Delta \to \infty$ the relative efficiency of the SPTE approached 1 from below. The relative efficiency of the SPTE is larger for smaller values of α near $\Delta = 0$ than for larger values of α for all σ and $d \neq 1$.

Since the prior non-sample information comes from previous knowledge/studies and, or, expert understanding of the phenomenon under investigation, the value of d is likely to be close to 0 and the value of Δ should not be much away from 0. In such a situation, the SPTE of the intercept vector has the largest relative efficiency and hence over performs the SRE and UE.

6 An example

To demonstrate the application of the method, we consider a data set on a health study from Plank (2004, p.8.31). The study investigates the systolic blood pressure of a group of patients divided in to the smoking and non-smoking categories. In the sample there are 10 smokers and 11 non-smokers. The age of the patients is the explanatory variable, X, and is divided in to X_1, the age of the smoking patients and X_2, that of the non-smoking patients. The systolic blood pressure is the response variable, Y. Regression lines of Y on X_1 and Y on X_2 have been fitted to the data for the two group of patients separately. The scatterplot and the fitted regression lines are given in Figure 3. The fitted regression lines for the two groups of data are

$$\hat{y}_1 = -21.9487 + 3.0911x_1, \ (R_1^2 = 0.9512) \tag{6.1}$$

$$\hat{y}_2 = 47.7437 + 1.6978x_2, \ (R_2^2 = 0.6761). \tag{6.2}$$

Other statistics relevant to the current study are $n_1Q_1 = 208.5$, $n_2Q_2 = 259.64$, $nQ = 468.14$ and $\hat{\beta} = 2.3184$, estimated slope from the combined sample. The observed value of the test statistic is 5.555 with a P-value of 0.0307. Hence there is not enough sample evidence to reject the null hypothesis of equal slopes, and thus the slopes of the two regression lines are not significantly different from one another for any $\alpha > 0.0307$. The graphs in Figure 3 represent the scatterplot and fitted regression lines of the data set for two different categories of respondents.

7 Concluding remarks

In this paper we have defined three different estimators for the intercept parameter of the two suspected parallel regression models. The performances of the three different estimators of the intercept parameters have been analyzed by using the criteria of

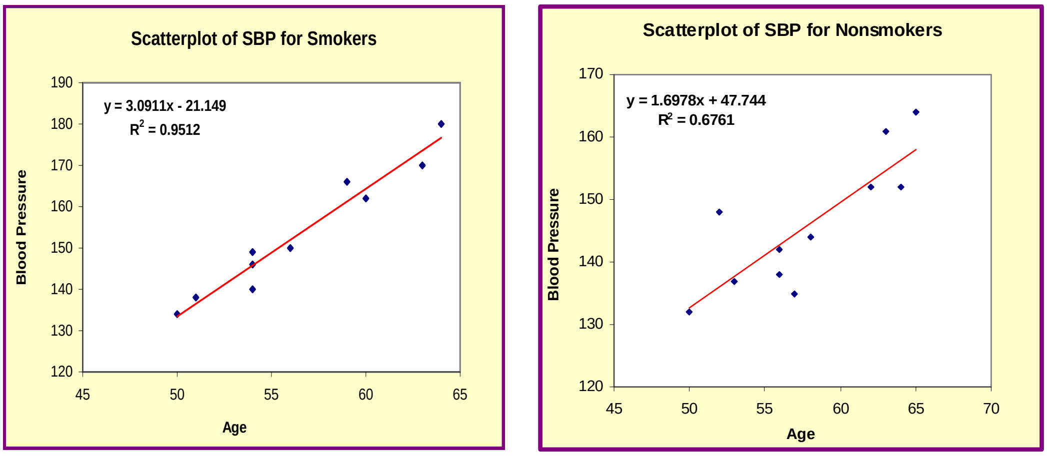


Figure 3: **Linear regression of SBP on Age for Smokers and Nonsmokers.**

quadratic bias and risk under quadratic loss. The SPTE has always smaller quadratic bias than the SRE, except at $\Delta = 0$. But the quadratic bias of the UE is always 0 for all values of Δ. Based on the criterion of quadratic bias, the UE is the best among the three estimators. However, the SPTE is the best among the bised estimators. Based on the quadratic risk criterion, the superiority of estimators depends on various conditions discussed in section 5 and the graphs displayed in Figure 2. The RE is the best only if $\Delta = 0$. In the face of uncertainty on the value of Δ, if Δ is likely to be small then the SPTE is the preferred option, regardless of the choice of α. One may use the UE as the best option if Δ is likely to be moderate, for which the quadratic risk of the SPTE reaches its maximum. For very large values of Δ the SPTE performs as good as the UE under the quadratic risk criterion. The source and nature of the non-sample prior information lead to believe that the values of both Δ and d are likely to be close to 0. In such a situation, the SPTE of the intercept vector has the largest relative efficiency and hence over performs the SRE and UE.

We have provided the marginal analysis of the problem. The joint study of the parameter sets of slopes and intercepts remains to be an open problem. Moreover, Stein-type shrinkage estimation is also possible for a set of $p > 2$ parallel regression models.

Acknowledgements

The author thankfully acknowledges some valuable suggestions from Professor A K Md E Saleh, Carleton University, Canada that helped improve an earlier version of the paper.

BIBLIOGRAPHY

Akaike, H. (1972). Information theory and an extention of maximum likelihood principle. Problems of Control and Informationn Theory. Academiai Kiado, Hungarian Academy of Science, 202-212.

Ali, A.M. and Saleh, A.K.Md.E. (1990). Estimation of the mean vector of a multivariate normal distribution under symmetry. *Jou. Statist. Comp. Simul.*, **35**, 209-226.

Bancroft, T.A. (1944). On biases in estimation due to the use of preliminary tests of significance. *Ann. Math. Statist.*, **15**, 190-204.

Bancroft, T.A. (1964). Analysis and inference for incompletely specified models involving the use preliminary test(s) of significance. *Biometrics*, **20**, 427-442.

Bancroft, T.A. (1972). Some recent advances in inference procedures using preliminary tests of significance. In *Statistical Papers* in honour of G.W. Snedecor, Iowa State Univ. Press, 19-30.

Bhoj, D.S. and Ahsanullah, M. (1993). Estimation of conditional mean in a linear regression model. *Biometrical Journal*, **35**, 791-799.

Bhoj, D.S. and Ahsanullah, M. (1994). Estimation of a conditional mean in a regression model after a preliminary test on regression coeficient . *Biometrical Journal*, **36**, 153-163.

Han, C.P. and Bancroft, T.A. (1968). On pooling means when variance is unknown. *Jou. Amer. Statist. Assoc.*, **29**, 21-34.

Hirano, K. (1977). Estimation with procedures based on preliminary test, shrinkage technique and information criterion, *Ann. Inst. Statist. Math.*, Vol. 1, 361-379.

Judge, G.G. and M.E. Bock (1978). The statistical implication of pre-test and Stein-rule estimators in econometrics. North-Holland, New York.

Khan, S. (2003). Estimation of the Parameters of two Parallel Regression Lines Under Uncertain Prior Information, Biometrical Journal, 44, 73-90.

Khan, S. and A.K.Md.E. Saleh (2001). On the comparison of the pre-test and Stein-type estimators for the univariate normal mean, Statistical Papers. 42(4). 451-473.

Khan, S. and A.K.Md.E. Saleh (1997). Shrinkage pre-test estimator of the intercept parameter for a regression model with multivariate Student-t errors. *Biometrical Journal*, Vol. 29, 131-147.

Khan, S. and A.K.Md.E. Saleh (1995). Preliminary test estimators of the mean for sampling from some Student-t populations. *Journal of Statistical Research*, Vol. 29, 67-88.

Kitagawa, T. (1963). Estimation after preliminary tests of significance. University of California Publications in Statistics, **3**, 147-186.

Mahdi, T.N., Ahmad, S.E. and Ahsanullah, M. (1998). Improved predicitons: Pooling two identical regression lines. *Jou. Appld. Statist. Sc.*, **7**, 63-86.

Plank, A.W. (2004). Experimental Design, Study Book, DEC, University of Southern Queensland, Australia

Puri, M.L. and Sen, P.K. (1973). Nonparametric Methods in Multivariate Analysis. Wiley, New York.

Saleh, A.K.Md.E. and Han, C.P. (1990). Shrinkage estimation in regression analysis. *Estadistica,* **42**, 40-63.

Saleh, A.K.Md.E. and Sen, P.K. (1978). Non-parametric estimation of location parameter after a preliminary test on regression. *Annals of Statistics,* **6**, 154-168.

Sen, P.K and Saleh, A.K.Md.E. (1985). On some shrinkage estimators of multivariate location. *Annals of Statistics.* **13**, 272-281.